DOCUMENT RESUME

ED 055 067 SP 005 310

AUTHOR Wardrop, James L.

TITLE Was New Century Teaching Students the Gates-MacGinitie Reading Tests?

INSTITUTION Illinois Univ., Urbana. Center for Instructional Research and Curriculum Evaluation.

SPONS AGENCY American Federation of Teachers, Washington, D.C.

PUB DATE [71]

NOTE 10p.

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Educational Responsibility; *Performance Contracts; Performance Criteria; *Program Effectiveness; *Testing Problems; *Test Wiseness

ABSTRACT

"Teaching the test" has been defined in terms of teaching those particular content knowledges or skills needed to answer the test items correctly. Evidence of several sorts examined in this paper clearly indicates that New Century was teaching students in Providence, R.I., the Gates-MacGinitie Reading Test, which was used to assess their vocabulary achievement. The coincidence between vocabulary taught in the instructional package and the vocabulary required to respond correctly to test items on the Gates-MacGinitie was determined to be much greater than could be attributed to chance, and the data showed that the teaching program needed be only moderately effective to improve substantially student gains in grade-equivalent scores in the test. On the basis of the analyses summarized in the paper, if the instructional materials are only 30 percent effective, scores should average nearly twice those which would normally be found. (MBM)

OCT 22 1971

WAS NEW CENTURY TEACHING STUDENTS THE GATES-MACGINITIE READING TESTS?

James L. Wardrop, Ph.D.
Center for Instructional Research and Curriculum Evaluation
University of Illinois at Urbana-Champaign

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 055067

I. Background

A student's performance on a standardized test is useful to us in making statements about his level of achievement only insofar as the test items have some congruence with our expectations about what a student should know at any particular time. For many reasons, vocabulary subtests provide a better vehicle for discussing this issue than do most others. Consider the findings that students at the end of second grade typically have a sight vocabulary of 1,000-1,100 words. Obviously, we could ascertain the extent to which a particular student meets this standard if we presented him with all the words which experts have agreed upon as rightly being included in this domain. Such an assessment procedure would enable us to make fairly precise statements about the student's level of accomplishment. However, the demands--both physical and psychological--on students, teachers, and testers of such a procedure would be unreasonable.

As an alternative, then, we sample from the vocabulary items in this domain some smaller number to which we ask students to respond. (Of course, some of the items in this vocabulary do not lend themselves to "testing" in any convenient way; e.g., "a," "the," "or," "I.") In the process of developing a standardized achievement test for vocabulary, many items are considered for inclusion; a smaller number are

*An opinion prepared at the request of the American Federation of Teachers, AFL-CIO.

SP 005 310

actually used in preliminary versions of the test; and a still smaller number appear in the final, published form(s) of the test. Those items which are retained are those which contribute most to the overall reliability and validity of the test.

The items which actually appear in a standardized test, then, are but a sample of the items which a student might reasonably be expected to know and on which he might reasonably be tested. It is on the basis of a student's performance on this sample of appropriate behaviors that we make inferences about his level of achievement in the domain of interest (vocabulary, in this case). Insofar as performance on the test may be considered as <u>representative</u> of what the student might be expected to do when exposed to that larger collection of behavior samples from which the test items were selected, that test performance is a valid indicator of his achievement level.

When the instructional process is such that the <u>particular</u> knowledges or skills required for successful performance on the <u>particular</u> test form(s) to be utilized are in fact specifically taught, the behaviors sampled in the test are no longer representative of the domain to which we wish to generalize. Thus, the most crucial consideration in whether "teaching the test" has occurred is whether the instructional content is of such a form as to render the test--and consequently normative inferences based on the test performance--

invalid as an indicator of the general body of knowledge to which inferences are to be made.*

It is on this basis, then, that the possibility that the contractor, New Century, has violated that provision of the contract in which the contractor "agrees that it will not teach the Gates-MacGinitie reading test..." (Section 20 of the contract agreement) should be evaluated.

II. Analysis

A student's vocabulary achievement is assessed, on any version of the Gates-MacGinitie Reading Test: Vocabulary, on the basis of his responses to anywhere from 48 (Level B, Grade 2) to 52 (Level C, Grade 3) items. The general form of these items is most easily represented in terms of what psychologists call a "paired-associates" task. That is, given a "stimulus" word (such as "incredible"), the student must associate with it some "response" word (such as "unbelievable"). As an illustration, consider item 30 from Primary C, Form 1 of the Gates-MacGinitie:

---

*"Preparing" students for a test can take several forms: providing them with practice in the test-taking situation by giving them experience with the item forms (but not the content) they will encounter on the test, providing them exposure to the specific content which they will encounter on the test, giving them experience with both the content and form of the test, and coaching them on the specific items from the test in the form in which they actually appear. The first of these is a legitimate form or preparation in that it tends to reduce the contribution of extraneous, situationally linked factors which are irrelevant to achievement in the domain of interest but which might affect performance on the test. The remaining three procedures are illegitimate (in that they invalidate the test as a representative sample of the behavior domain to which inferences are to be made), with the last being the most blatant and dishonest attempt to invalidate the test and inappropriately enhance student performance.

30. medicine

meadow
iron
spider
drug

Here, the student is presented with the stimulus word "medicine." The correct response is "drug." If "teaching the test" has occurred, we would expect to find this same pairing of words in Unit 5 of *Word Wizards*. In fact, item 30 of part 6 of Unit 5 of *Word Wizards* is:

30. A drug you take when you're sick is

a. brook
b. medicine
c. hurry

In this instance, then, the same words are paired in the instructional materials as are used in assessing student vocabulary on the Gates-MacGinitie Reading Test.

Granting that some common paired associates appearing in the Gates-MacGinitie will appear in an instructional package such as *Word Wizards* simply by chance (and the contractor is not to be penalized for such "chance coincidence"), the question remains: Is the coincidence between vocabulary taught in *Word Wizards* and vocabulary required to respond correctly to test items on the Gates-MacGinitie attributable to chance?

If a definition of "coincidence" based on the "paired-associate" conceptualization described above is adopted, the results presented in Table 1 are obtained. Without reference to any other evidence, these data are conclusive:

- - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - -

TABLE 1

Analysis of the Relationship of Instructional Content (*Word Wizards*) to Test Items (Gates-MacGinitie Reading Test: Vocabulary)

Grade	Test Level	Test Form	No. of Items	*Word Wizards* Unit	No. of Items[a]	No. of Items Common[b]
2	B	1	48[c]	R	120	31
		2	48[c]			31
3	C	1	12+40[d]	S	120	3+20[d]
		2	12+40			3+20
4-6	D	1	50	T	120	26
		2	50			26
		3	50			0
7-9	E	2	50	U	120	29

[a]Number of stimulus words taught in the Unit.

[b]Items in which the "correct" stimulus-response pairs from the test appear in *Word Wizards*.

[c]The items in Level B are picture-stimulus/word-response items. At this level, the correspondence of test items and instructional content was assessed in terms of whether or not the meanings of the correct response words for the test were presented in Unit R of *Word Wizards*.

[d]Form C-1 and Form C-2 each contain 12 picture/word items (like those in Level B) and 40 word-stimulus/word-response items of the type illustrated in the "medicine/drug" example in the text. For Form C-1, three of the 12 picture/word items and 20 of the 40 word/word items occur in *Word Wizards*; for Form C-2, the corresponding numbers are 5 and 20.

1. The commonality for every test form examined except D-3 is much greater than would be expected by chance (especially at the upper grade levels, where the child's vocabulary should be comprised of several thousand words*).

2. Additionally, the same degree of overlap occurs for both forms at grade two (B-1 and B-2), for the word/word items on both forms at grade three (C-1 and C-2), and for two of the three forms at grades four through six (D-1 and D-2).

3. Equally telling is the complete lack of commonality between Unit T of *Word Wizards* and test form D-3.

Supplementary data provided by Judith F. Barry in her July 13th memorandum to Dr. Bernardo and Mr. Kramer support the findings presented above. Not only did she find an extremely high overlap between *Word Wizards* and forms of the Gates-MacGinitie but, in addition, her analysis, while proceeding on a slightly different basis from the one reported here, also provides comparative data relating *Word Wizards* to other standardized reading tests. In general, her figures indicate a commonality of less than twenty percent for most tests other than the Gates-MacGinitie. The Barry memorandum also highlights the relatively uncommon words (based on standard lists of common words) which occur both on the Gates-MacGinitie tests and in *Word Wizards*.

The conclusion is inescapable: New Century, through its *Word Wizards* materials, was teaching the word associations required to respond correctly to items on the Gates-MacGinitie Reading Test: Vocabulary. The coincidence between teaching materials and test items cannot be attributed to "chance." If the *Word Wizards* program was at

*Even at the end of grade two, the typical sight vocabulary is 1,000 words or more.

all effective, the results from the Gates-MacGinitie have been invalidated as indicants of vocabulary achievement for contract-program pupils.

In particular, what are the possible effects of exposure to the content of *Word Wizards* on the change in grade-equivalent scores for students in the contract program? One approach to this question is in terms of "typical" students of various sorts. For example, consider a second grader who was performing at an "average" level upon entry into the program in December. He would obtain a raw score of 25-26 on one of the forms of Level B of the Gates-MacGinitie (Vocabulary), giving him a grade equivalent score (GE) of 2.3 years. If he remains an "average" student at the time of the exit test in May, he will obtain a GE score of 2.9 (a raw score of-34). Thus, this student will have gained 8-9 items and 0.6 years. Assume that one wishes this student to show a gain of at least 1.0 years GE, to a terminal level of 3.3 years. For Level B of the Gates-MacGinitie, a GE score of 3.3 years corresponds to a raw score of 36 so improving the student's performance by only <u>two</u> items will result in a GE increase of 0.4 years.

To relate this potential improvement to the possible effects of exposure to *Word Wizards* is the next task. If a typical second-grade student will in fact correctly answer 34 of the 48 vocabulary items from either form of Level B of the Gates-MacGinitie without any special intervention, there are only 14 items remaining on which he could show further improvement. If we assume that the proportion of those 14 items included in *Word Wizards* is the same as the proportion of all 48 items appearing in *Word Wizards* (31/48 or .65), then the student will have been exposed

to nine of the items he would ordinarily have missed. He needs only recall two of those nine items in order to improve his apparent vocabulary growth during the contract period from 0.6 years GE to 1.0 years GE. These figures represent an effectiveness rate for the *Word Wizards* program of only 22 percent.

- - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - -

Similar arguments could be developed for "typical" students in grades three through eight and for other student groups in all grades. Table 2 presents just such an analysis. In addition to "average" students (those at the fiftieth percentile in the norm distribution), students ranking at the sixteenth (one standard deviation below the norm group mean), the thirty-first (one-half standard deviation below the norm group mean), and the sixty-ninth (one-half standard deviation above the norm group mean) percentiles have also been included in this Table. On the basis of the analyses summarized in Table 2, one can conclude that, if the instructional materials (the *Word Wizards* program) are only thirty percent effective on the average, student gains in grade-equivalent scores should average nearly twice the magnitude of those which would ordinarily be found.

III. Synopsis

"Teaching the test" has been defined in terms of teaching those particular content knowledges or skills needed to answer the test items correctly. Converging evidence of several sorts clearly indicates that New Century was teaching the Gates-MacGinitie Reading Test. An

TABLE 2

Analysis of Possible Impact of Exposure to *Word Wizards* on Grade-Equivalent Scores on the Gates-MacGinitie Reading Test: Vocabulary for Selected Subgroups of Students

Grade	Percentile Rank	"Normal" Performance Lecember	May	Additional Items for 1.0 Year's Gain	Percent in *Word Wizards*	Needed Effectiveness Rate
2	16	15	21	7	65	41
	31	20	27	5	65	39
	50	26	34	2	65	22
	69	33	41	–	65	0
3	16	19	23	6	46	46
	31	25	29	6	46	55
	50	30	35	3	46	37
	69	35	40	1	46	20
4	16	14	17	5	52	29
	31	18	22	3	52	21
	50	22	26	3	52	25
	69	27	30	2	52	27
5	16	19	21	6	52	55
	31	24	26	4	52	33
	50	29	31	2	52	22
	69	33	35	1	52	14
6	16	23	26	4	52	31
	31	28	30	3	52	30
	50	33	34–35	1–2	52	25
	69	35–36	38	1	52	17
7	16	13	14	3	58	15
	31	16–17	17–18	2	58	11
	50	20	21	2	58	12
	69	24	25	3	58	21
8	16	15–16	16–17	2	58	11
	31	19	20	2	58	11
	50	23	24	3	58	20
	69	27	28	1	58	8

analysis of the normative data for the test shows that the teaching program (*Word Wizards*) need be only moderately effective to improve student gains in grade-equivalent scores on that test substantially.